Measuring physical performance via self-report in healthy young adults

Kuijer, W.; Gerrits, E.H.; Reneman, M.F.

Published in:
Journal of Occupational Rehabilitation

Publication date:
2004

Journal of Occupational Rehabilitation, Vol. 14, No. 1, March 2004 (© 2004)

Measuring Physical Performance via Self-Report in Healthy Young Adults

W. Kuijer,[1,2,3] E. H. J. Gerrits,[1] and M. F. Reneman[2]

Discrepancies exist in literature as to what extent self-reporting can replace performance-based testing. To answer this question, self-reports and performance tests should measure identical constructs. Previous studies did not measure identical constructs. The objective of our study was to investigate to what extent self-reporting can replace performance-based testing. Seventy-two healthy subjects were tested. The constructs of the self-reports and the performance tests covered the same components to enable a comparison of self-reports and performance test results. Three different self-reports and a performance test were used to measure physical performance. Additionally, rating of perceived exertion was measured after the subjects lifted a reference weight to predict maximal lifting performance. The controls were age, gender, educational level, subject's participation in fitness, availability of reference data, motivation, attitude, general self-efficacy, and mood. Results showed that all lifting tasks could be predicted, though not solely via self-reporting. A prediction of the performance test results with a margin of ±5 kg of error could be made for at least 79% of the subjects, via gender, self-reporting, and subject's participation in fitness. Self-reporting may not replace performance testing, although performance testing can be predicted with a margin of error of ±5 kg for at least 79% of the healthy subjects.

KEY WORDS: task performance and analysis; self-assessment; exertion.

INTRODUCTION

Physical performance can be estimated via different kinds of instruments. Examples of instruments used to estimate physical performance are self-reports, proxy-reports, performance testing, clinical observation, or a combination of these instruments (1). Estimation of physical performance is often used in clinical practice to determine someone's ability to work. Physicians appear to rely strongly on a patient self-report. Self-report is one's verbal or written estimation of one's capacity to perform activities. Performance-based tests usually referred to as functional capacity evaluations (FCEs), measure performance of work-related

[1]Institute of Human Movement Sciences, University of Groningen, Haren, The Netherlands.
[2]University Hospital Groningen, Center for Rehabilitation, Haren, The Netherlands.
[3]Correspondence should be directed to W. Kuijer, MSc, University Hospital Groningen, Department of Rehabilitation Medicine, Hanzeplein 1, P.O. Box 30.001, 9700 RB Groningen, The Netherlands; e-mail: w.kuijer@rev.azg.nl.

activities (2). FCEs are time-consuming and expensive, while self-reports are less expensive and more practical to use. The question is to what extent physical performance can be estimated via self-reporting.

Some suggest that self-reporting can be used to replace performance-based testing (3), at least for screening healthy subjects (1). Moderate to strong correlations between measured physical performances and self-report of function have been reported (4,5). Others support the idea that physical performance cannot be measured via self-reporting (6–9) and that direct observation, as opposed to self-reporting, is the better indicator of patient behavior (10). This latter premise may be questioned however. The studies mentioned did compare not only self-reports and performance-based tests, but also differences in construct, context, and item scaling (9,11). To answer the question to what extent physical performance can be estimated via self-reporting, self-reports and performance tests should measure identical constructs. The validity of the questionnaire that were used previously is expected to be weak because such questionnaires do not measure the construct of physical performance.

To estimate physical performance, the predicting behavior should be described as accurately as possible (12). Subjects must understand what kind of behavior is required and what conditions the questionnaires apply to (13). Accessibility of information in memory and contextual cues appears to be of key importance (14). Motivational and cognitive factors can confound the assessment of "true capacity" (15). Additionally, motivation and emotion are known to influence the perception during physical performance as well as the performance itself (16). Furthermore, a relationship exists between self-efficacy and the degree of effort a subject expends on a test (13,17–19). Important predictors of general behavior for healthy subjects and therefore also for performance during an FCE and self-reported performance are attitude and self-efficacy (19,20). Lackner *et al.* (19) also showed that functional self-efficacy was significantly related to behavioral measures of physical function. Efficacy expectations alone will not however produce the desired performance if the component abilities are missing (13). Performance testing preceding self-reporting makes the subjects more aware of their true physical capabilities, but further research is needed into ways to combine self-reporting and performance measures (21).

In this study, different self-reports were compared with performance test results, that is, different lifting tasks. Differences in outcome between performance test results and self-reports were described and explained. The goal of this study was to determine to what extent self-reporting can be used to replace performance-based testing. Information from performance-based tests is often used to determine one's ability to work (22). Therefore answering this question contributes to the development of a cost-effective method to determine ability to work, it does however not directly assess its validity.

METHODS

Subjects

A convenience sample of 72 healthy subjects (36 male and 36 female) participated in this study. All but one were students. Their mean age was 22 years (ranging from 19 to 28 years). Before participating, all subjects declared in an informed consent that they were healthy and agreed to participate voluntarily and at their own risk.

Procedures

Firstly, personal data (age, gender, and educational level) were obtained via a self-constructed questionnaire, as were data of lifting experience and sporting activities. One's attitude towards self-reporting and performance testing was measured with a self-constructed 10-cm Visual Analogue Scale (VAS). Secondly, self-efficacy was measured via the ALCOS, short form (SF) (ALgemene COmpetentie Schaal, a Dutch version of the General self-efficacy scale) (23). Thirdly, subjects were asked to fill out three different self-reports to measure self-estimated physical performance, that is, the maximal amount of weight they can lift. Fourthly, the subjects performed four different lifting tasks, with 6-min rests between each task. After the self-report measures and after each lifting task, the subject's motivation was measured with a self-constructed 10-cm VAS, and mood was measured using a Profile Of Mood States questionnaire (POMS-SF, translated version into Dutch) (24). The total test duration was approximately 2 h 15 min.

Measures

Self-Reports

Three questionnaires were constructed to estimate physical performance. Each type of self-report added progressively more information. Self-report 1 consisted of open-ended questions about the maximum amount of weight they could lift; self-report 2 consisted of closed questions using everyday examples as a reference; self-report 3 consisted of asking subjects to lift a reference weight, and then asking what percentage of his or her maximum performance the weight represented. These latter execution-related questions resembled the performance test most closely. Each self-report covered the same components as the lifting tasks, to equalize the constructs of the questionnaires and the performances; in each self-report, lifting heights (lifting from waist to shoulder or from waist to overhead and back), and the amounts of repetitions were reported, dependent on the relevant lifting task. Additionally, illustrations concerning the relevant lifting task accompanied each question. Perceived exertion was measured after lifting the reference weight, using a Rating of Perceived Exertion scale (RPE scale) (16). An example of an open-ended question was: "What is the maximum weight you can lift one time from your waist to overhead and back?" An example of a closed question was: "Can you lift one kilo (pack of sugar) from your waist to overhead and back?" The questions in this self-report ranged from 1 to 40 kg for women and from 1 to 60 kg for men, with 5-kg increments at each question. For the execution-related questions, the reference weight the subjects had to lift was determined by estimating 70% of average maximum lifting ability expected in their age group (based on results of a pilot study). Seventy percent of maximal effort was used because the highest test–retest correlation coefficients on RPE-scale were obtained for 70% effort of maximum (16).

Performance Tests

Subjects lifted a starting weight from waist to shoulder or waist to overhead and back for a set of one or five repetitions. Heart rate was measured after each set, and weight was added until a maximum was reached, following the procedures used in different functional capacity evaluation protocols. Maximum lifting performance was determined when a strength maximum or a maximal acceptable heart rate ((220−age)*85%) was reached, or safety was no longer guaranteed. Subjects were also allowed to discontinue the test themselves. In this case, no safe maximal performance could be determined. The lifting tasks were derived from two FCEs. Task 1, lifting from waist to shoulder and back, one repetition, is the Upper Lifting Strength test from the Ergo-Kit FCE.[4] Task 2, lifting from waist to shoulder and back, five repetitions, was derived from the previous lifting task. Task 3, lifting from waist to overhead and back, five repetitions, was modified from the Isernhagen Work System FCE (IWS FCE).[5] Task 4, lifting from waist to overhead and back, one repetition, was derived from the previous lifting task. A lab situation was established with the materials needed for the lifting tasks. Materials needed for the tasks included a measurement frame with two adjustable shelves (2.5-cm increments), a stopwatch, a heart rate monitor, a plastic box, and weights.

For lifting from waist to shoulder and back, one repetition, unpublished data showed good test–retest reliability for a workers compensation population (25). Because maximum performance was required in this study, the Ergo-Kit termination criteria were equalized to the IWS termination criteria. Lifting from waist to overhead and back, five repetitions, has demonstrated good test–retest reliability for healthy subjects (unpublished data, 26), and for lower back pain patients (27,28). Good intra- and inter rater reliability was established in healthy subjects (2). No reliability data were available for the other lifting tasks.

Additional Variables

Self-efficacy was measured with the ALCOS-SF (23), measuring the subjects' expectations of their capacities in general. This questionnaire consists of 17 questions with response possibilities on a 5-point Likert scale. The sum score ranges from 100 to 500. The reliability and construct validity of the scale is satisfactory (29). Motivation and attitude were measured with a self-constructed 10-cm VAS. The scale ranges from "not motivated" to "very motivated" and from "very negative" to "very positive." Mood was measured using the POMS-SF (translated version into Dutch) (24). It measures depression, anger, fatigue, vigor, and tension on a 5-point Likert scale. The reliability of the POMS-SF is satisfactory (30) and is an excellent alternative to the original POMS (31). Little validity research is available for the POMS-SF.

Data Analysis

Strong over- and underestimations–defined as estimations of the subject's test results under or over three standard deviations (SDs) of the mean performance test results—were removed from analyses list-wise. This resulted in removal of a maximum of seven estimations per test, depending on the lifting task and/or self-report. Specifically, lifting from waist to shoulder, one repetition, $n = 7$; lifting from waist to shoulder, five repetitions, $n = 3$;

[4]Ergo-Kit Functional Capacity Evaluation, Ergo Control B.V., Buurserstraat 214, 7544 RG Enschede, The Netherlands.

[5]Isernhagen Work System Functional Capacity Evaluation, 1015 E. Superior Street Duluth, Minnesota 55802.

lifting from waist to overhead, five repetitions, $n = 5$; lifting from waist to overhead, one repetition, $n = 6$. Descriptive statistics were calculated for performance test results and self-reports and for absolute differences between performance test results and matched self-report. The percentages of maximal lifting ability for the execution-related questions were calculated. Spearman's rank correlation coefficients were calculated between performance test results and self-reports. Multiple linear regression analyses were used to predict performance test results via self-reporting, controlling for age, gender, educational level, attitude, general self-efficacy, availability of reference data, subject's participation in fitness, motivation, mood, and rating of perceived exertion. Predictors found significant in these analyses were entered in subsequent multiple linear regression analyses to establish a model for the prediction of performance test results. In this second regression analysis, not all of the above-mentioned subjects were removed from the analyses, because of the list-wise removal, additionally, one or two subjects were excluded because of missing data in the variables controlled for. Including all 72 subjects, percentages were described of predictions of performance test results with a margin of error of ±5 kg, using the unstandarized residual within the (second) regression analysis. All analyses were performed using the Statistical Package for Social Sciences (SPSS 10.1 for windows).

Interpretation of Data

An alpha value of 0.05 was used to determine statistical significance. Correlation coefficients must be higher than 0.75 to be relevant in a clinical situation (criterion for concurrent validity) (32). Criteria for the linear regression were probability of F to enter ≤ 0.050, and probability of F to remove ≥ 0.100. Within the regression analyses, predictors were deemed significant if $\alpha \leq 0.05$. To decide whether self-reporting can replace performance-based testing, the prediction of an individual test result should fall within a margin of error of ±5 kg, of the actual individual test result; the unstandarized residual should not exceed ±5 kg. This is based on a criterion of the Dutch government, which uses 5-kg increments in its assessment methods to indicate clinically important differences.

RESULTS

Means, SDs, and minimal and maximal values of performance test results and self-reports are presented in Table I. This table represents the values of the self-report scores and test results, applied to the different lifting tasks. Means, SDs, and ranges of differences between performance test results and self-reports are presented in Table II. These differences between performance test results and self-reports are absolute differences, and therefore do not have to be equal to the differences presented in Table I. Cell 1 represents the absolute mean difference between the open-ended questions and lifting from waist to shoulder and back, one repetition, which is 10 kg with an SD of 6.6 kg and a range of 26 kg.

The estimated amount of 70% of maximal lifting ability (which the subjects had to lift for the execution-related questions) turned out to range between 40 and 135%. The average percentage of maximal lifting ability for the different lifting tasks ranged between 60 and 80%.

Table I. Descriptive Statistics of Test Results and Self-Reports in Kilograms

	Open-ended questions			Closed questions			Execution-related questions			Performance test result		
	M	SD	Min–Max	*M*	SD	Min–Max	*M*	SD	Min–Max	*M*	SD	Min–Max
Lifting from waist to shoulder, one repetition	23.6	12.6	5–60	31.2	10.5	10–55	37.0	12.4	20–83	30.3	7.7	20–50
Lifting from waist to shoulder, five repetitions	18.5	12.5	3–60	17.5	10.9	1–50	27.2	9.4	15–57	24.5	6.1	12–42
Lifting from waist to overhead, five repetitions	15.0	11.1	2–50	12.8	8.4	1–40	22.4	7.1	13–44	19.2	5.3	8–30
Lifting from waist to overhead, one repetition	19.2	11.4	3–50	26.1	10.5	5–55	30.0	8.7	16–47	24.8	6.7	14–38

Table II. Absolute Differences Between Performance Test Results and Self-Reports in Kilograms

	Open-ended questions			Closed questions			Execution-related questions		
	M	SD	Range	*M*	SD	Range	*M*	SD	Range
Lifting from waist to shoulder, one repetition	10.0	6.6	26.0	6.2	4.5	20.0	7.1	8.3	32.3
Lifting from waist to shoulder, five repetitions	10.0	5.4	22.0	9.7	6.3	22.5	6.9	6.9	27.5
Lifting from waist to overhead, five repetitions	8.7	5.6	25.0	7.9	5.1	21.0	6.9	8.7	60.9
Lifting from waist to overhead, one repetition	9.2	5.4	22.0	6.2	5.4	30.0	5.9	7.9	49.0

Spearman's rank correlation coefficients were calculated between performance test results and self-reports. All correlation coefficients were significant at the level of 0.001. Correlation coefficients are presented in Table III.

Performing multiple regression analyses, results showed that age, educational level, availability of reference data, motivation, attitude, general self-efficacy, mood, open-ended questions, and execution-related questions did not significantly contribute to the prediction of performance test results (all p values were ≥ 0.05, data not shown). Subsequent regression analyses showed that lifting from waist to shoulder and back, one repetition, was predicted by the closed questions and gender ($F(2, 65) = 104.2$, $p \leq 0.001$); lifting from waist to shoulder and back, five repetitions, was predicted by gender ($F(1, 70) = 67.5$, $p \leq 0.001$); lifting from waist to overhead and back, five repetitions, was predicted by gender and subject's participation in fitness ($F(2, 68) = 46.7$, $p \leq 0.001$); lifting from waist to shoulder and back, one repetition, was predicted by the closed questions, gender, and subject's participation in fitness ($F(3, 65) = 74.8$, $p \leq 0.001$). These variables accounted for 76, 48, 57, and 77% of the adjusted variance, respectively. Results of this (second) regression analysis (variables entered which significantly contributed to the prediction of performance test results) are presented in Table IV.

For 81% of the subjects the individual test result was predictable for lifting from waist to shoulder and back, one repetition. For 79% of the subjects the individual test result was predicable for lifting from waist to shoulder and back, five repetitions. For 80% of the subjects the individual test result was predicable for lifting from waist to overhead and back, five repetitions. For 84% of the subjects the individual test result was predicable for lifting from waist to shoulder and back, one repetition. All predictions included a margin of error of ± 5 kg (Table V).

Table III. Correlation Coefficients (r) Between Test Results and Self-Reports

	Open-ended questions	Closed questions	Execution-related questions
Lifting from waist to shoulder, one repetition	0.55	0.69	0.43
Lifting from waist to shoulder, five repetitions	0.50	0.55	0.48
Lifting from waist to overhead, five repetitions	0.55	0.57	0.51
Lifting from waist to overhead, one repetition	0.56	0.72	0.52

Note. All correlations were significant ($p < 0.001$).

Table IV. Multiple Regression Analyses Predicting Physical Performance

Performance test and significant predictors	ß	Adjusted R^2 for model	df	F value equation
Lifting from waist to shoulder, one repetition		0.76	65	104.2**
Constant	18.46**			
Gender	8.98**			
Closed questions	0.23**			
Lifting from waist to shoulder, five repetitions		0.48	70	67.5**
Constant	20.21**			
Gender	8.54**			
Lifting from waist to overhead, five repetitions		0.57	68	46.7**
Constant	14.78**			
Gender	7.67**			
Subject's participation in fitness	2.45*			
Lifting from waist to overhead, one repetition		0.77	65	74.8**
Constant	14.92**			
Gender	8.34**			
Closed questions	0.20**			
Subject's participation in fitness	2.07*			

Note. Variables entered that significantly contributed to the prediction of performance test results.
*$p < 0.05$. **$p < 0.001$.

DISCUSSION

The goal of this study was to determine to what extent self-reporting can be used to replace performance-based testing. Results showed that correlation coefficients between self-reports and performance tests were too low to be relevant in a clinical situation. Additionally, results from the linear regression showed that all lifting tasks could be predicted, though not solely via self-reporting. A prediction of the performance test result with a margin of error of ±5 kg could be made for at least 79% of the subjects. In conclusion, self-reporting may not replace performance testing, although performance testing can be predicted with a margin of error of ±5 kg for at least 79% of the healthy subjects, via gender, self-reporting, and/or subject's participation in fitness.

Although some previous literature showed that physical performance can be assessed by RPE (Borg, 1982a; Borg and Ottoson, 1986, in 16), and significant correlations were found between RPE and amount of weight patients were able to lift (correlation coefficients unknown) (33), other literature confirmed our findings. Perceptual differences were found when different treadmill protocols were used (34). Performance of a postural tolerance test was only weakly associated with perceived exertion for an elevated work test ($r = 0.23$) and for a forward bending test ($r = 0.23$) (7). In conclusion, RPE scales should not be used to predict maximal performance.

Table V. Prediction of Performance Test Results

	With a margin of error of ±5 kg		Exceeding ±5 kg	
	n	%	n	%
Lifting from waist to shoulder, one repetition ($n = 68$)	59	81.9	9	12.5
Lifting from waist to shoulder, five repetitions ($n = 72$)	57	79.2	15	20.8
Lifting from waist to overhead, five repetitions ($n = 71$)	58	80.6	13	18.1
Lifting from waist to overhead, one repetition ($n = 69$)	61	84.7	8	11.1

An explanation of the significant contribution of subject's participation in fitness to the prediction of lifting from waist to overhead and back, one and five repetitions, and not to the prediction of the other tests, could be that this specific lifting task of the Isernhagen Work Systems FCE is more comparable with performing fitness activities than are the lifting tasks derived from the Ergo-Kit FCE. Subjects performing the lifting task from waist to overhead and back benefit from participating in fitness. Therefore, fitness participation can significantly contribute to the prediction of lifting from waist to overhead and back.

Previous studies showed that functional self-efficacy was strongly related to performance testing, suggesting that self-efficacy would contribute significantly to the prediction of performance test results. In our study, self-efficacy did not significantly contribute to the prediction of performance test results. It should be mentioned that in our study general self-efficacy was measured, instead of functional self-efficacy. Additionally, in our study healthy subjects were tested instead of patients, and little interindividual variability in self-efficacy scores was found. This may explain the absence of a significant contribution.

In this study, reliable conclusions for groups could be drawn because 72 healthy subjects were studied. A gradual construction was made to measure performance. It was expected that the execution-related questions provided the best information for the subject and, consequently, enabled the subject to make the most accurate estimation of physical performance. This self-report was not however the self-report that provided the best information for lifting ability. In the lifting tasks, asking closed questions was the best self-report to estimate physical performance.

Healthy subjects were tested; they varied little in age, motivation, attitude, mood, and self-efficacy. This was not a representative group of subjects, because performance testing is usually used for measuring the performance of work-related activities in patients. When estimating physical performance of patients instead of healthy subjects, and to determine if the results found in this study can be generalized to a group of patients, other controlling variables should be included. These variables may be depression (15,17,35), kinesiofobia (15,17), pain (36), pain behavior (3), pain self-efficacy (1,15), functional self-efficacy (19), functional status (36), outcome expectations (1,15), harm-belief (3), and disability compensation (15,37). These variables may influence performance during an FCE, as well as self-reported performance.

After performing this study, it is possible to answer the question to what extent physical performance can be estimated via self-reporting in healthy subjects, because self-reports and performance tests measured identical constructs. Self-reporting may not replace the performance tests lifting from waist to shoulder and waist to overhead and back, for a set of one or five repetitions. However, results of performance tests can be predicted with a margin of error of ± 5 kg for at least 79% of the healthy subjects, via gender, self-reporting and/ or subject's participation in fitness. It should be mentioned that the margin of error of ± 5 kg in predicting performance test results could result in an over- or underestimation of 26% in this group of healthy subjects. Individuals can exceed this percentage of 26%. It is necessary to determine whether a margin of error of ± 5 kg is acceptable when using self-reporting in a clinical environment. Besides, it is unknown to what extent individual subjects vary in test performances and in estimations. If the natural variation of performance test results exceeds 5 kg, the 5 kg criterion will become inappropriate, because predictions over 5 kg may not by definition result in a wrong estimation of physical performance. Furthermore, strong over- and underestimations removed from analyses were defined as estimations of the subject's test results under or over three SDs of the mean performance test results. This definition was based on the idea that subjects with these estimations can be filtered out immediately, and should be tested anyhow. If this criterion was sharpened to two standard deviations of the mean performance test results, the prediction would probably be more accurate because of the smaller deviations.

To be able to answer the question to what extent self-reporting can be used when estimating physical performance in individual patients (i.e., in disability determination), further research is needed to determine natural variation in performance testing and estimating, and to combine self-reporting and performance-based testing. Although information gathered from performance testing is often used to determine ability to work, this study has not investigated to what extent this way of self-reporting can assess the ability to work. More research is needed to answer the question whether or not estimating physical performance can predict ability to work in patients, as it is used for this goal.

ACKNOWLEDGMENTS

The authors thank Jan Plat for placing testing materials at our disposal, and for providing the courses to become a registered test leader (WK). Additionally, we thank Stefan IJmker, who helped collect data.

REFERENCES

1. Wunderlich GS. *Measuring functional capacity and work requirements, summary of a workshop.* Washington, DC: National Academy Press, 1999.
2. Reneman MF, Jaegers SMHJ, Westmaas M, Goëken LNH. The reliability of determining effort level of lifting and carrying in a functional capacity evaluation. *Work* 2002; 18(1): 23–28.
3. Jensen MP, Romano JM, Turner JA, Good AB, Wald LH. Patient beliefs predict patient functioning: Further support for a cognitive–behavioral model of chronic pain. *Pain* 1999; 81: 95–104.
4. Simmonds MJ, Olson SL, Jones S, Hussein T, Lee CE, Novy D, Radwan H. Psychometric characteristics and clinical usefulness of physical performance tests in patients with low back pain. *Spine* 1998; 23 (22): 2412–2421.
5. Simonsick EM, Kasper JD, Guralnik JM, Bandeen-Roche K, Ferrucci L, Hirsch R, Leveille S, Rantanen T, Fried LP. Severity of upper and lower extremity functional limitation: Scale development and validation with self-report and performance-based measures of physical function. *J Gerontol B Psychol Sci Soc Sci* 2001; 56B (1): S10–S19.
6. Daltroy LH, Phillips CB, Eaton HM, Larson MG, Partridge AJ, Logigian M, Liang MH. Objectively measuring physical ability in elderly persons: The physical capacity evaluation. *Am J Public Health* 1995; 85 (4): 558–560.
7. Reneman MF, Bults MMWE, Engberts LH, Mulders KKG, Goëken LNH. Measuring maximum holding times and perception of static elevated work and forward bending in healthy young adults. *J Occup Rehabil* 2001; 11 (2): 87–97
8. Piela CR, Hallenberg KK, Geoghegan AE, Monsein MR, Lindgren BR. Prediction of functional capacities. *Work* 1996; 6: 161–164.
9. Wijlhuizen GJ, Ooijendijk W. Measuring disability, the agreement between self-evaluation and observation of performance. *Disabil Rehabil* 1999; 21 (2): 61–67.
10. Catalano MD, Campagna TDG, Chiappetta CN, Peters S, Dale L. The inter-rater reliability and validity of self-report of risk factors for cumulative trauma disorders within the hand therapist population—A pilot study. *Work* 1999; 12 (2): 189–194.
11. Reneman MF, Jorritsma W, Schellekens JMH, Goëken LNH. Concurrent validity of questionnaire and performance based disability measurements in patients with chronic low back pain. *J Occup Rehabil* 2002; 12 (3): 119–130.
12. Azjen I. From intentions to actions: A theory of planned behavior. In: Kuhl, J. & Beckman, J. eds. *Action control: From Cognition to behavior.* Berlin: Springer-Verlag, 1985, pp. 11–39.

13. Bandura A. Self-efficacy: Toward a unifying theory of behavioral change. *Psychol Rev* 1977; 84 (2): 191–215.
14. Stone AA, Turkkan JS, Bachrach AC, Jobe JB, Kurtzman HS, Cain VS. *The science of self-report. Implications for research and practice.* Mahwah, NJ: Erlbaum, 2000.
15. Watson PJ. Non-physiological determinants of physical performance in musculoskeletal pain. *Syllabus IASP refresher courses on pain management held in conjunction with the 9*th *World Congress on pain.* Vienna, Austria. Aug. 22–27, 1999.
16. Borg G. *Borg's perceived exertion and pain scales.* Champaign, IL: Human Kinetics, 1998.
17. Kaplan GM, Wurtele SK, Gillis D. Maximal effort during functional capacity evaluation: An examination of psychological factors. *Arch Phys Med Rehabil* 1996; 77: 161–164.
18. Matheson LN, Matheson ML, Grant J. Development of a measure of perceived functional ability. *J Occup Rehabil* 1993; 3 (1): 15–29.
19. Lackner JM, Carosella AM, Feuerstein M. Pain expectancies, pain, and functional self-efficacy expectancies as determinants of disability in patients with chronic low back disorders. *J Consult Clin Psychol* 1996; 64 (1): 212–220.
20. Brug J, Schaalma H, Kok G, Meerten RM, Molen van der HT. *Gezondheidsvoorlichting en gedragsverandering, een planmatige aanpak.* Assen, The Netherlands: Van Gorcum & Comp. B.V, 2000.
21. Daltroy LH, Larson MG, Eaton HM, Phillips CB, Liang MH. Discrepancies between self-reported and observed physical function in the elderly: The influence of response shift and other factors. *Soc Sci Med* 1999; 48: 1549–1561.
22. Matheson LN, Isernhagen SJ, Hart DL. Relationships among lifting ability, grip force, and return to work. *Phys Ther* 2002; 82 (3): 249–256.
23. Scherer M, Maddux JE, Mercandante B, Prentice-Dunn S, Jacobs B, Rogers RW. The self-efficacy scale: Construction and validation. *Psychol Rep* 1982; 51: 663–671.
24. Cluydts RJG. *Gemoedstoestanden en slaap*. Proefschrift, Brussel: Vrije Universiteit Brussel, 1979.
25. Schellekens J. *Reconditionering als warming-up voor reïntegratie: tussenrapport* [Reconditioning as a warm-up for reintegration]. Groningen, The Netherlands: Arbeidspsychologie, 2001.
26. Dijkink A, Kuis M, Reneman MF, Brouwer S, Goeken LNH. *Onderzoek naar de betrouwbaarheid van de Functionele Capaciteit Evaluatie van Isernhagen* [Study of the reliability of the Isernhagen Functional Capacity Evaluation]. Groningen The Netherlands: Leeronderzoek Instituut voor Bewegingswetenschappen, 2000.
27. Gross DP, Battié NC. Reliability of safe maximal lifting determination of a functional capacity evaluation. *Phys Ther* 2002; 82 (4): 364–371.
28. Reneman MF, Dijkstra PU, Westmaas M, Goëken LNH. Test–retest reliability of lifting and carrying in a 2-day functional capacity evaluation. *J Occup Rehabil* 2002; 12 (4): 269–275.
29. Bosscher RJ, Smit JH, Kempen GIJM. Algemene competentie verwachtingen bij ouderen: een onderzoek naar de psychometrische kenmerken van de Algemene competentieschaal (ALCOS)[Global expectations of self-efficacy in the elderly: An investigation of psychometric characteristics of the General Self-efficacy Scale] *Ned Tijdschr Psychol* 1997; 52: 239–248.
30. Wald FDM, Mellenbergh GJ. De verkorte versie van de Nederlandse vertaling van de Profile of Mood States (POMS). *Ned tijdschr Psychol* 1990; 45: 86–90.
31. Curran SL, Andrykowski MA, Studts JL. Short form of the profile of mood states (POMS-SF): Psychometric information. *Psychol Assess* 1995; 7 (1): 80–83.
32. Innes E, Straker L. Validity of work-related assessments. *Work* 1999; 13 (2): 125–152.
33. King ML, Dracup KA, Woo MA. Predictors of isotonic exercise in patients with heart failure. *Med Sci Sports Exerc* 2001; 33 (7): 1090–1095. [Abstract].
34. Whaley MH, Brubaker PH, Kaminsky LA, Miller CR. Validity of rating of perceived exertion during graded exercise testing in apparently healthy adults and cardiac patients. *J Cardiopulm Rehabil* 1997; 17: 261–267.
35. Fylkesnes K, Forde OH. The Tromso Study: Predictors of self-evaluated health—Has society adopted the expanded health concept? *Soc Sci Med* 1991; 32 (2): 141–146.
36. Fishbain DA, Abdel-Moty A, Cutler R, Khalil TM, Sadek S, Rosomoff RS, Rosomoff HL. Measuring residual functional capacity in chronic low back pain patients based on the dictionary of occupational titles. *Spine* 1994; 19 (8): 872–880.
37. Fylkesnes K, Forde OH. Determinants and dimensions involved in self-evaluation of health. *Soc Sci Med* 1992; 35 (3): 271–279.